TEST
DOI 10.1007/s11749-014-0396-0

ORIGINAL PAPER

Random assignment processes: strong law of large numbers and De Finetti theorem

Ricardo Vélez · Tomás Prieto-Rumeau

Received: 28 November 2013 / Accepted: 27 July 2014

Abstract In the framework of a random assignment process—which randomly assigns an index within a finite set of labels to the points of an arbitrary set—we study sufficient conditions for a strong law of large numbers and a De Finetti theorem. In particular, this yields a family of finite-valued nonexchangeable random variables that are conditionally independent given some other random variable, that is, they verify a De Finetti theorem. We show an application of the De Finetti theorem and the law of large numbers to an estimation problem.

Keywords Random assignment processes · Exchangeability · Strong laws of large numbers · De Finetti theorem

Mathematics Subject Classification 60G09

1 Introduction

In the probability literature, a De Finetti theorem usually refers to a result ensuring that some random variables are conditionally independent given some other random variable. In this sense, the classical De Finetti theorem states that an infinite number of exchangeable random variables $\{Y_i\}_{i \in I}$ taking values in $\{0, 1\}$ are conditionally independent (and so, they are conditionally i.i.d.); see, e.g., Feller (1966), Fristedt

This research was supported by the Spanish Ministerio de Economía y Competitividad, Grant Number MTM2012-31393.

R. Vélez · T. Prieto-Rumeau (✉)
Statistics Deparment, UNED, Madrid, Spain
e-mail: tprieto@ccia.uned.es

R. Vélez
e-mail: rvelez@ccia.uned.es

Published online: 06 September 2014

and Gray (1997). Here, exchangeability means invariance of the distribution under finite permutations of the terms; so that any finite vector $(Y_{i_1}, \ldots, Y_{i_n})$ has the same distribution, which depends only on n. De Finetti (1930) proved that there is a random variable ξ with values in $[0, 1]$ such that, conditional on ξ, the Y_i are i.i.d. random variables with $\xi = P\{Y_i = 1 \mid \xi\}$.

Generalizations of this "first" De Finetti theorem were aimed towards showing that exchangeable random variables taking values in domains L larger that $\{0, 1\}$ are conditionally independent as well. De Finetti (1937) has given a proof for real-valued random variables, ξ being now an arbitrary distribution on $\mathbb{R}$. Later, Hewitt and Savage (1955) proved the result when L is an arbitrary compact Hausdorff space. The case of a Borel space L (isomorphic to some Borel subset of $[0, 1]$ endowed with its Borel sigma-field) is studied in Fristedt and Gray (1997, Chapter 27) and Kallenberg (1997, Chapter 8). However, Dubins and Freedman (1979) give an example showing that this theorem might fail in a separable metric space.

Finite families of exchangeable random variables have also been studied, since they are related to sampling without replacement from a finite population. More precisely, ξ is then replaced by the (random) empirical distribution and, conditional on ξ, $(Y_1, \ldots, Y_n)$ arise from sampling without replacement from the support of ξ. Diaconis and Freedman (1980a) deal with this case and analyze the total variation distance between the distribution of $(Y_1, \ldots, Y_n)$ and the closest mixture of i.i.d. random variables, as a method for proving the De Finetti theorem. See also Kerns and Székely (2006).

De Finetti's type results have also evolved in many different directions. For instance, Diaconis and Freedman (1980b) show that mixtures of recurrent Markov chains, obtained by randomly selecting the transition matrix, are characterized by "partial exchangeability", meaning that two sets of trajectories beginning with the same number of transitions between each couple of states have the same probability. Continuous-time extensions of such results are given by Freedman (1996), not only for Markov chains but also for stationary independent increments' processes.

The case of random arrays $\{X_{i,j}\}_{i,j\in\mathbb{N}}$ satisfying exchangeability properties under separate or joint permutations of rows and columns has been independently analyzed by Aldous (1981, 1982) and Hoover (1982), and it was simplified and generalized by Kallenberg (1989, 1992). Dovbysh and Sudakov (1982) consider the case of symmetric positive definite random matrices invariant under permutations of their rows and columns. See also Panchenko (2010).

Exchangeable partitions of a set of natural numbers were considered by Kingman (1978), motivated by some problems in population genetics. Later, Gnedin (1997), Gnedin and Pitman (2004), and Pitman (1995) pursued this topic.

Exchangeable random measures μ, defined as those for which the random variables $\mu(B_1), \ldots, \mu(B_n)$ are exchangeable whenever $\lambda(B_1) = \cdots = \lambda(B_n)$ for a given measure λ, were first studied by Kallenberg (1973, 1975) and the results were collected in the monograph Kallenberg (1976). Closed random subsets $M \subset \mathbb{R}$ with Lebesgue measure zero, whose complement is a disjoint union of open intervals (a_i, b_i), are called exchangeable if, for each $\epsilon > 0$ and conditional on the number N_ϵ of intervals of length at least ϵ, their lengths $b_i - a_i$ are an exchangeable sequence. Their study has been considered in Kallenberg (1982a, b).

For more details and other issues, two specific monographs have been written by Aldous (1985) and, more recently, by Kallenberg (2005).

As can be seen, De Finetti's theorems are mostly related to some kind of exchangeability or symmetry adapted to the context of the problem, which implies that, conditional on a random selection of some parameters, the involved random variables have not only independence properties, but also, as a consequence of the symmetry, some homogeneity in their distributions (in most cases they are identically distributed).

Our recent research, however, has focused on obtaining De Finetti's theorems without exchangeability assumptions. In this sense, sufficient conditions for a family $\{Y(x)\}_{x\in E}$ of non-identically distributed $\{0, 1\}$-valued random variables to verify a De Finetti theorem are proposed in Vélez and Prieto-Rumeau (2009, 2011) by introducing the concept of a (weighted) random selection process, with $Y(x) = 1$ when the element $x \in E$ is "selected", while $Y(x) = 0$ if $x \in E$ is "rejected". Then, under adequate hypotheses, the random variables $\{Y(x)\}_{x\in E}$ are shown to be conditionally independent, though they are neither i.d. nor exchangeable (this $\{0, 1\}$ case is related to a short paper by Pitman 1978). In Vélez and Prieto-Rumeau (2010) we have proved that these hypotheses are also necessary to get a De Finetti theorem for such random selection processes.

In Vélez and Prieto-Rumeau (2008), a generalization of the random selection processes was proposed, named random assignment processes. The random variables $\{Y(x)\}_{x\in E}$ are now supposed to take values in a finite set $L = \{0, 1, \ldots, \ell\}$. So, the value of $Y(x)$ can be interpreted as the label or index assigned to the point $x \in E$. The labels are assumed to be randomly assigned according to a multidimensional weight function (the details are given in Sect. 2 below). Then, under further appropriate conditions, it is shown that the L-valued random variables $\{Y(x)\}_{x\in E}$ satisfy a De Finetti theorem, that is, they are conditionally independent though, once again, they are neither i.d. nor exchangeable. In the present paper, the sufficient conditions for a De Finetti theorem proposed in Vélez and Prieto-Rumeau (2008) are widely generalized, and in particular, we give a satisfactory answer to the conjecture stated in the conclusions of Vélez and Prieto-Rumeau (2008). In addition, a strong law of large numbers (related to the random variable with respect to which the $\{Y(x)\}_{x\in E}$ are conditionally independent) is studied.

In the literature there is an exception to the fact that the De Finetti theorem is related to identically distributed random variables. This has to do with the relations between the De Finetti theorem and sufficiency, as initially explored by Dynkin (1978) and Lauritzen (1982); a key result in this sense is given in Diaconis and Freedman (1984, Theorem 1.1) that is also included in Aldous (1985, Section 18). The corresponding setting is a sequence of random variables $(X_1, \ldots, X_n, \ldots)$ such that a sufficient statistic T_n exists for each $n \geq 1$ and the conditional distribution of $(X_1, \ldots, X_n)$ given $T_n = t$ is $Q_n(t, \cdot)$, which must satisfy some compatibility assumptions. The class M of distributions P of the sequence for which this statement is true, called partially exchangeable in Diaconis and Freedman (1984), is a convex set and there is a set A in

$$\Sigma = \bigcap_n \sigma(T_n, X_{n+1}, X_{n+2}, \ldots)$$

with $P(A) = 1$ such that, in A, $Q_n(T_n, \cdot)$ converges to a probability measure $Q \in M$; these Q are trivial in Σ and give the extremes of M. Any other $P \in M$ is a mixture of these Q with respect to the restriction of P to Σ. However, this result does not ensure the independence of the random variables under the extremes of M; so that the theorem encompasses as well cases where the conditional independence of De Finetti's theorem is granted with others where it fails.

In this context, our paper will show that the probability measure associated to a random assignment process is partially exchangeable, and gives conditions ensuring that the extreme distributions are multinomial-like distributions, so that the conditional independence of the variables indeed holds. The corresponding mixing probabilities are then supported on the multidimensional unit simplex.

The rest of the paper is organized as follows. In Sect. 2 we give a formal definition of a random assignment process and state our main results, whose proofs are postponed to later sections. Namely, Sect. 3 proves several useful preliminary results, while the proofs of the strong law of large numbers and the De Finetti theorem are given in Sect. 4. An application of the results in this paper to an estimation problem is studied in Sect. 5. Our conclusions are stated in Sect. 6.

2 Random assignment processes

Most of the material in this section is borrowed from Vélez and Prieto-Rumeau (2008). For clarity of exposition, however, we will state the corresponding results.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, with expectation operator $\mathbb{E}$, let E be an arbitrary set, and consider the set of indices $L = \{0, 1, 2, \ldots, \ell\}$ for some positive integer ℓ. Let $\boldsymbol{Y} = \{Y(x)\}_{x \in E}$ be a family of random variables

$$Y(x) : \Omega \to L \quad \text{for every } x \in E.$$

There are no restrictions on the nature of the set E. It may be helpful, however, to think of E as a subset of an Euclidean space. In this way, the random variable $Y(x)$ assigns a color in L to the point $x \in E$. Actually, we will see that the most challenging case is when E is a denumerable set; surprisingly, our results are easier to obtain for an uncountable set E. In the sequel, we will impose some conditions—see (C1) and (C2) below—on the probability measure $\mathbb{P}$ so that $\boldsymbol{Y}$ has some suitable properties.

2.1 Notation

Let $\{\boldsymbol{e}_i\}_{1 \le i \le \ell}$ be the canonical basis of $\mathbb{R}^\ell$, and let $\boldsymbol{e_0} = \boldsymbol{0} \in \mathbb{R}^\ell$. The random variables $Y(x)$ may be alternatively represented by means of the random vectors

$$N(x) = \boldsymbol{e}_{Y(x)}.$$

Letting $\mathcal{B}$ be the ring of all finite subsets of E, we can define for each $B \in \mathcal{B}$

$$N(B) = \sum_{x \in B} N(x) = (N_1(B), \dots, N_\ell(B))$$

which takes values in

$$R(\#B) = \left\{ \boldsymbol{r} \in \mathbb{N}^\ell : \overline{\boldsymbol{r}} \le \#B \right\},$$

where $\#B$ denotes the cardinal of the set B and $\overline{\boldsymbol{r}} = \sum_{i=1}^{\ell} r_i$ is the sum of the components of $\boldsymbol{r} \in \mathbb{N}^\ell$. We let $N_0(B) = \#B - \sum_{i=1}^{\ell} N_i(B)$, and accordingly, we define also $r_0 = \#B - \overline{\boldsymbol{r}}$. Consequently, $N(B) = \boldsymbol{r}$ means that there are $N_i(B) = r_i$ points x in B with $Y(x) = i$ (say, of color i) for each $0 \le i \le \ell$.

2.2 Invariant σ-algebras

On L we will consider the discrete σ-algebra (consisting of all the subsets of L). On the set of all functions from E to L, denoted by L^E, we will consider the product σ-algebra; see Kallenberg (1997, p. 2).

Given a finite set $B \in \mathcal{B}$, let $\mathcal{G}_B$ be the sub-σ-algebra of $\mathcal{F}$ consisting of all the events of the form $\{\boldsymbol{Y} \in F\}$, where F is a measurable subset of L^E which is invariant under any permutation of the elements of B. Following the interpretation in terms of coloring, it should be evident that a family of colorings of the points in E is invariant under permutations of the elements of B if and only if the colors outside B are fixed, while the frequencies of each color inside B remain also fixed. In fact:

Lemma 1 *Given $B \in \mathcal{B}$, $\mathcal{G}_B$ is the smallest σ-algebra for which $N(B)$ and $Y(x)$, for $x \in B^c$, are measurable, denoted by $\sigma(N(B), \{Y(x)\}_{x \in B^c})$.*

For an arbitrary subset $A \subseteq E$, we define the σ-algebra of events invariant under any finite permutation in A as

$$\mathcal{G}_A = \bigcap_{B \in \mathcal{B}, B \subseteq A} \mathcal{G}_B. \tag{2.1}$$

Obviously, if $A_1 \subseteq A_2$ then $\mathcal{G}_{A_2} \subseteq \mathcal{G}_{A_1}$.

2.3 Definition of a random assignment process

Let $\boldsymbol{m} : E \to (0, \infty)^\ell$ be an arbitrary function. We will refer to

$$\boldsymbol{m}(x) = (m_1(x), \dots, m_\ell(x))$$

as the vector of weights of $x \in E$. Later, we will impose additional conditions on $\boldsymbol{m}$. Given a finite set $B \in \mathcal{B}$, we will use the notation $m_i(B) = \sum_{x \in B} m_i(x)$ for $1 \le i \le \ell$.

Definition 1 A family of L-valued random variables $\boldsymbol{Y} = \{Y(x)\}_{x\in E}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, or equivalently, the family $\boldsymbol{N} = \{N(B)\}_{B\in\mathcal{B}}$, is a random assignment process on E with weight function $\boldsymbol{m}$ (that will be also referred to as a random assignment process on $(E, \boldsymbol{m})$) if the following conditions hold for each $B \in \mathcal{B}$:

(C1) Conditional on $N(B)$, the random variables $\{Y(x)\}_{x\in B}$ and $\{Y(x)\}_{x\in B^c}$ are independent.[1]
(C2) For all $i = 1, \ldots, \ell$ and $x \in B$,

$$\mathbb{P}\{Y(x) = i \mid N(B) = \boldsymbol{e}_i\} = \frac{m_i(x)}{m_i(B)}.$$

The condition (C1) means that, once the numbers of the different indices attributed to the points inside B are known, their positions in B do not depend on the indices attributed to the points outside B; this definition has a Markovian-like flavor. On the other hand, the condition (C2) establishes that if a unique positive index $i \in L$ is to be located in B, then this positive index corresponds to $x \in B$ with probability proportional to its weight $m_i(x)$.

It is important to note that one of the basic issues that we shall discuss is the existence itself of a random assignment process for given E and $\boldsymbol{m}$.

2.4 Conditional distributions

It is easy to prove Vélez and Prieto-Rumeau (2008, Proposition 2.1) that for any $B \in \mathcal{B}$, $\boldsymbol{r} \in R(\#B)$, and any disjoint sets $B_1, \ldots, B_\ell \subset B$ with $\#B_i = r_i$ for $i = 1, \ldots, \ell$, we have

$$\mathbb{P}\left\{\bigcap_{i=1}^{\ell}\bigcap_{x\in B_i}\{Y(x) = i\} \mid N(B) = \boldsymbol{r}\right\} = \frac{1}{S(B, \boldsymbol{r})}\prod_{x\in B_1} m_1(x)\cdots\prod_{x\in B_l} m_\ell(x), \quad (2.2)$$

where the normalizing factor $S(B, \boldsymbol{r})$ is the sum of the products in the righthand side of (2.2) over all the disjoint subsets of B with $\#B_i = r_i$ for $i = 1, \ldots, \ell$. In case that $\boldsymbol{r} \notin R(\#B)$, we define $S(B, \boldsymbol{r}) = 0$.

Fix $B \in \mathcal{B}$ and a subset $C \subset B$. As a direct consequence of (2.2) (cf. Vélez and Prieto-Rumeau 2008, Proposition 2.3), for any $\boldsymbol{r} \in R(\#B)$ and $\boldsymbol{k} \in R(\#C)$ such that $\boldsymbol{r} - \boldsymbol{k} \in R(\#B - \#C)$, the probability that $N(C) = \boldsymbol{k}$ conditional on $N(B) = \boldsymbol{r}$ is the following "generalized" multidimensional hypergeometric distribution

$$\mathbb{P}\{N(C) = \boldsymbol{k} \mid N(B) = \boldsymbol{r}\} = \frac{S(C, \boldsymbol{k})\, S(B - C, \boldsymbol{r} - \boldsymbol{k})}{S(B, \boldsymbol{r})}. \quad (2.3)$$

[1] For a definition of conditional independence see Kallenberg (1997, p.109).

More generally, if $C_1, \ldots, C_s$ are disjoint subsets of $B \in \mathcal{B}$, then

$$\begin{aligned}\mathbb{P}\{N(C_1) = \boldsymbol{k}_1, \ldots, N(C_s) = \boldsymbol{k}_s \,|\, N(B) = \boldsymbol{r}\} \\ = \frac{S(C_0, \boldsymbol{k}_0) S(C_1, \boldsymbol{k}_1) \cdots S(C_s, \boldsymbol{k}_s)}{S(B, \boldsymbol{r})}\end{aligned} \tag{2.4}$$

with $C_0 = B - \cup_{t=1}^{s} C_t$ and $\boldsymbol{k}_0 = \boldsymbol{r} - \sum_{t=1}^{s} \boldsymbol{k}_t$.

2.5 Marginal distributions

The above results show that the distribution of the random assignment process $\boldsymbol{N}$ inside each $B \in \mathcal{B}$ is determined by the marginal distribution of $N(B)$. Hence, the distribution of the random assignment process is completely determined by a family of compatible marginal distributions for each $N(B)$ with $B \in \mathcal{B}$, meaning that for every disjoint B and C in $\mathcal{B}$ the total probability rule

$$\mathbb{P}\{N(B) = \boldsymbol{k}\} = \sum_{\boldsymbol{r} \in R(\#C)} \mathbb{P}\{N(B \cup C) = \boldsymbol{k} + \boldsymbol{r}\} \frac{S(C, \boldsymbol{r}) S(B, \boldsymbol{k})}{S(B \cup C, \boldsymbol{k} + \boldsymbol{r})} \tag{2.5}$$

holds.

Clearly, if E is finite, then we can choose an arbitrary distribution for $N(E)$, and then derive the joint distributions of the random assignment process using (2.4) with $B = E$. Therefore, to avoid this trivial situation, in what follows we will assume that the set E is not finite.

As shown in Vélez and Prieto-Rumeau (2008, Proposition 2.5), an obvious choice of compatible marginal distributions for any E and $\boldsymbol{m}$ is the following mixture of "generalized" multinomial distributions: for $B \in \mathcal{B}$

$$\mathbb{P}\{N(B) = \boldsymbol{k}\} = \int_{\Delta_\ell} \frac{S(B, \boldsymbol{k})\, \xi_1^{k_1} \cdots \xi_\ell^{k_\ell} \xi_0^{\#B - \bar{k}}}{\prod_{x \in B} \boldsymbol{\xi} \cdot \boldsymbol{m}(x)}\, F(d\boldsymbol{\xi}) \quad \text{for all } \boldsymbol{k} \in R(\#B), \tag{2.6}$$

where $\Delta_\ell = \{\boldsymbol{\xi} \in [0, 1]^L : \boldsymbol{\xi} \cdot \mathbf{1} = \xi_0 + \cdots + \xi_\ell = 1\}$ and F is an arbitrary distribution on Δ_ℓ. We note that, to simplify some expressions, in (2.6) we let

$$m_0(x) = 1 \quad \text{and} \quad \boldsymbol{m}(x) = \big(m_0(x), m_1(x), \ldots, m_\ell(x)\big) \quad \text{for all } x \in E.$$

Remark 1 Let us now comment on the particular case $m_i(x) \equiv 1$ for all $i \in L$. In this case, $S(B, \boldsymbol{r})$ equals the multinomial coefficient, that is,

$$S(B, \boldsymbol{r}) = \binom{\#B}{r_0, r_1, \ldots, r_\ell} = \frac{\#B!}{r_0! r_1! \cdots r_\ell!}$$

and the random variables $Y(x)$ are exchangeable; see Remark 2.4 in Vélez and Prieto-Rumeau (2008). This also gives the intuition for the generalized multinomial distribution in (2.6). Obviously, the total probability rule (2.5) holds when the marginal distributions are multinomial.

Remark 2 In the previous setting, the index $0 \in L$ seems to play a particular role as a "default" index. However, if m is a function from E to $(0, \infty)^{\ell+1}$, with

$$\boldsymbol{m}(x) = (m_0(x), m_1(x), \ldots, m_\ell(x)) \quad \text{for } x \in E,$$

then, replacing (C2) with

(C2)' For all $i, j \in L$ with $i \neq j$,

$$\mathbb{P}\{Y(x) = i \mid N_i(B) = 1, N_j(B) = \#B - 1\} = \frac{m_i(x)/m_j(x)}{\sum_{y \in B} m_i(y)/m_j(y)}$$

we would reach identical results, with the factor $\prod_{x \in B_0} m_0(x)$ inserted in the righthand side of (2.2).

Thus, there is no loss of generality when letting $j = 0$ and $m_0(x) \equiv 1$ in (C2)'. This explains why, in the sequel, the index 0 has no special relevance whereas, on the contrary, the quotients $m_i(x)/m_j(x)$ will be particularly relevant for our analysis.

In case that (2.6) holds for some F, the random variables $\{Y(x)\}_{x \in E}$ are conditionally independent given $\boldsymbol{\xi}$, with conditional distributions

$$\mathbb{P}\{Y(x) = i \mid \boldsymbol{\xi}\} = \mathbb{P}\{N(x) = \boldsymbol{e}_i \mid \boldsymbol{\xi}\} = \frac{\xi_i\, m_i(x)}{\boldsymbol{\xi} \cdot \boldsymbol{m}(x)} \quad \text{for } i \in L. \tag{2.7}$$

Therefore, we say that a De Finetti theorem holds.

2.6 Main results

Our goal in this paper is to give conditions on the weight function $\boldsymbol{m}$ ensuring that, given a random assignment process on $(E, \boldsymbol{m})$, the marginal distributions $N(B)$ for $B \in \mathcal{B}$ are necessarily as in (2.6) for some distribution F on Δ_ℓ (and so, a De Finetti theorem holds).

It was shown in Vélez and Prieto-Rumeau (2008, Theorem 3.5) that a De Finetti theorem holds if the following condition is satisfied:

Condition (A) There exists a sequence[2] $A = \{x_n\}_{n \geq 1} \subset E$ such that

$$0 < \lim_{n \to \infty} m_i(x_n) < \infty \quad \text{for every } 1 \leq i \leq \ell.$$

Letting $A_n = \{x_q\}_{1 \leq q \leq n}$, the condition (A) above implies that the $m_i(A_n)$ grow to infinity, as $n \to \infty$, at the same speed (proportional to n) for all $i \in L$. With respect to the strong law of large numbers, it is proved in Vélez and Prieto-Rumeau (2008, Theorem 3.4) that as $n \to \infty$

$$\left(\frac{N_1(A_n)}{n}, \ldots, \frac{N_\ell(A_n)}{n} \right) \text{converges a.s. to a random vector } \boldsymbol{\xi} \in \Delta_\ell. \tag{2.8}$$

[2] A sequence $A = \{x_n\}_{n \geq 1} \subset E$ will always refer to a countable nonfinite subset of E.

It was also shown in Vélez and Prieto-Rumeau (2008, Example 3.8) that any random assignment process on an uncountable set E satisfies condition (A). Therefore, the challenging case is precisely when the set E is countable. Here, we will improve Vélez and Prieto-Rumeau (2008, Theorem 3.5) by replacing (A) with the weaker condition (A*) below, in which we will use the notation $m_{j|i}(x) = m_j(x)/m_i(x)$ for $x \in E$ and $i, j \in L$.

Condition (A*) There exists a sequence $A = \{x_n\}_{n\geq 1} \subset E$ such that, for every i and j in L, the limit $\lim_{n\to\infty} m_{j|i}(x_n) = m_{j|i}$ exists, with $0 \leq m_{j|i} \leq \infty$, and moreover,

$$m_{j|i}(A) = \sum_{n=1}^{\infty} m_{j|i}(x_n) = \infty.$$

We note that, regardless the condition (A*) holds, there is indeed a sequence $A \subset E$ such that the limits $m_{j|i}$ exist (just choose a suitable subsequence of any given sequence in E). Besides, the series $m_{j|i}(A)$ diverges whenever $m_{j|i} > 0$. Hence, the condition (A*) just imposes the existence of a sequence A such that $m_{j|i}(A) = \infty$ even if $m_{j|i} = 0$.

When (A) is replaced with (A*), the statement itself of the strong law of large numbers (2.8) becomes much more involved; see Theorem 2 below. To give an overall idea, let us mention that it establishes the existence of a random vector $\boldsymbol{\xi}$ taking values in Δ_ℓ such that, letting

$$\Omega_M = \big\{N_i(A) > 0 \ \ \forall\, i \in M \quad \text{and} \quad N_j(A) = 0 \ \ \forall\, j \notin M\big\}$$

for $M \subset L$, we have

$$\mathbb{P}\left(\bigcup_{M\subset L} \Omega_M\right) = 1,$$

and moreover, for each $M \subset L$ with $\mathbb{P}(\Omega_M) > 0$, there exists a set $C_0 \subset M$ such that $\mathbb{P}(\cdot \mid \Omega_M)$-a.s. the following limits hold as $n \to \infty$: for all $i \in C_0$

$$\frac{N_i(A_n)}{n} \longrightarrow \frac{\xi_i}{\alpha_0}, \text{ while } \frac{N_j(A_n)}{m_{j|i}(A_n)} \longrightarrow \frac{\xi_j}{\alpha_0} \quad \forall\, j \notin C_0,$$

where $\alpha_0 = \sum_{i\in C_0} \xi_i$. This strong law of large numbers allows to get the following De Finetti result.

Theorem 1 *Let $\boldsymbol{Y}$ be a random assignment process on $(E, \boldsymbol{m})$, and suppose that the condition (A*) is verified for some sequence $A \subset E$. Then there exists a $\mathcal{G}_A$-measurable random vector $\boldsymbol{\xi}$ with values in Δ_ℓ such that the random variables $\{Y(x)\}_{x\in E}$ are conditionally independent given $\boldsymbol{\xi}$ and (2.7) holds, that is,*

$$\mathbb{P}\{Y(x) = i \mid \boldsymbol{\xi}\} = \frac{\xi_i\, m_i(x)}{\boldsymbol{\xi}\cdot \boldsymbol{m}(x)} \quad \textit{for all } x \in E \textit{ and } i \in L.$$

Consequently, for all $B \in \mathcal{B}$ and $\boldsymbol{k} \in R(\#B)$,

$$\mathbb{P}\{N(B) = \boldsymbol{k}\} = \mathbb{E}\left[\frac{S(B, \boldsymbol{k})\, \xi_1^{k_1} \cdots \xi_\ell^{k_\ell} \xi_0^{\#B - \bar{k}}}{\prod_{x \in B} \boldsymbol{\xi} \cdot \boldsymbol{m}(x)}\right]. \tag{2.9}$$

This results generalize the classical De Finetti theorem for sequences of exchangeable finite-valued random variables. The next result gives insight into the structure of the σ-algebras $\mathcal{G}_A$ when (A*) holds.

Corollary 1 *Under the conditions of Theorem 1 we have $\mathcal{G}_A = \sigma(\boldsymbol{\xi})$ provided that both σ-algebras are completed in $\mathcal{F}$. More generally, for any $E' \subset E$ such that $A \subset E'$, we have $\mathcal{G}_{E'} = \sigma(\boldsymbol{\xi})$ when both σ-algebras are completed in $\mathcal{F}$.*

In Sect. 3 we will prove several useful preliminary results that give a considerable insight into the behavior of any random assignment process along any sequence $A \subset E$. The proofs of Theorem 1 and Corollary 1 are given in Sect. 4.

3 Preliminary results

Throughout this section, we will suppose that $A = \{x_n\}_{n \geq 1}$ is a sequence in E and will use the notation $A_n = \{x_1, \ldots, x_n\}$ for $n \geq 1$.

Lemma 2 *Let $\boldsymbol{Y}$ be a random assignment process on $(E, \boldsymbol{m})$, and let A be an arbitrary sequence in E. Given any $B, C \in \mathcal{B}$ with $C \subset B \cup A$, we have*

$$\lim_{n \to \infty} \mathbb{P}\{N(C) = \boldsymbol{k} \mid \mathcal{G}_{B \cup A_n}\} = \mathbb{P}\{N(C) = \boldsymbol{k} \mid \mathcal{G}_{B \cup A}\} \text{ a.s.}$$

for any $\boldsymbol{k} \in R(\#C)$. Moreover, if $C \subset B \cup A_n$ (for large enough n), it is

$$\mathbb{P}\{N(C) = \boldsymbol{k} \mid \mathcal{G}_{B \cup A_n}\} = S(C, \boldsymbol{k})\, \frac{S((B \cup A_n) - C, N(B \cup A_n) - \boldsymbol{k})}{S(B \cup A_n, N(B \cup A_n))}. \tag{3.1}$$

Proof Note that $\{\mathcal{G}_{B \cup A_n}\}_{n \geq 1}$ is a reverse filtration, and therefore, the sequence $\{\mathbb{P}\{N(C) = \boldsymbol{k} \mid \mathcal{G}_{B \cup A_n}\}\}_{n \geq 1}$ is a reverse martingale. The convergence

$$\mathbb{P}\{N(C) = \boldsymbol{k} \mid \mathcal{G}_{B \cup A_n}\} \longrightarrow \mathbb{P}\{N(C) = \boldsymbol{k} \mid \mathcal{G}_{B \cup A}\}$$

follows from the reverse martingale convergence theorem (see, e.g., Fristedt and Gray 1997, p. 484). Now, as a consequence of the condition (C1), if $C \subset B \cup A_n$ then $N(C)$ and $\{N(x_q)\}_{q > n}$ are conditionally independent given $N(B \cup A_n)$. Therefore,

$$\mathbb{P}\{N(C) = \boldsymbol{k} \mid \mathcal{G}_{B \cup A_n}\} = \mathbb{P}\{N(C) = \boldsymbol{k} \mid N(B \cup A_n)\}.$$

Finally, according to (2.3), we conclude that (3.1) holds. □

Lemma 2 above ensures that, for all $\boldsymbol{k} \in R(\#C)$, the limit

$$\lim_{n\to\infty} \frac{S((B \cup A_n) - C, N(B \cup A_n) - \boldsymbol{k})}{S(B \cup A_n, N(B \cup A_n))}$$

exists almost surely. It is also of particular interest to define, for each $x \in A$ and $i \in L$, $\eta_i(x) = \mathbb{P}\{Y(x) = i \mid \mathcal{G}_A\}$, and so, almost surely,

$$\eta_i(x) = \mathbb{P}\{N(x) = \boldsymbol{e}_i \mid \mathcal{G}_A\} = m_i(x) \lim_{n\to\infty} \frac{S(A_n - x, N(A_n) - \boldsymbol{e}_i)}{S(A_n, N(A_n))}. \tag{3.2}$$

Then, by letting

$$\xi_i(x) = \frac{\eta_i(x)/m_i(x)}{\sum_{j\in L} \eta_j(x)/m_j(x)},$$

we have that (3.2) yields

$$\frac{\xi_j(x)}{\xi_i(x)} = \lim_{n\to\infty} \frac{S(A_n - x, N(A_n) - \boldsymbol{e}_j)}{S(A_n - x, N(A_n) - \boldsymbol{e}_i)} \text{ a.s.} \tag{3.3}$$

for all $x \in A$ and $i, j \in L$. Moreover, the $\boldsymbol{\xi}(x) = \{\xi_i(x)\}_{i\in L}$ are random ($\mathcal{G}_A$-measurable) probability vectors (that is, taking values in Δ_ℓ) that allow to express

$$\eta_i(x) = \frac{\xi_i(x) \cdot m_i(x)}{\boldsymbol{\xi}(x) \boldsymbol{\cdot} \boldsymbol{m}\ (x)}.$$

At this point, note that if we could show that $\boldsymbol{\xi}(x)$ does not depend on x, then we would be close to (2.7). As a first step towards this result, we will now show that the $\{\xi_i(x)\}_{x\in A}$ are simultaneously positive or simultaneously zero, according to the presence or absence of the index i along the sequence A (that is, depending on whether $N_i(A) > 0$ or $N_i(A) = 0$).

Proposition 1 *Let* $\boldsymbol{Y}$ *be a random assignment process on* $(E, \boldsymbol{m})$ *and let* A *be any sequence in* E. *For all* $i \in L$ *we have (up to sets of probability* 0*)*

$$\{N_i(A) = 0\} = \{\xi_i(x) = 0,\ \forall\, x \in A\} = \{\xi_i(x_1) = 0\}. \tag{3.4}$$

(Clearly, $\xi_i(x)$ *may be replaced with* $\eta_i(x)$ *in* (3.4) *above.)*

Proof Let $\bar{\Omega}$ be a measurable subset of Ω for which, for all $x \in A$ and $i \in L$, $\eta_i(x)$ equals

$$m_i(x) \lim_{n\to\infty} \frac{S(A_n - x, N(A_n) - \boldsymbol{e}_i)}{S(A_n, N(A_n))}$$

(recall (3.2)) and $\sum_{i\in L} \eta_i(x) = 1$. According to Lemma 2, we have $\mathbb{P}(\bar{\Omega}) = 1$. In $\bar{\Omega}$, if $N_i(A) = 0$ then $\eta_i(x) = 0$ for all $x \in A$, since $S(A_n - x, N(A_n) - \boldsymbol{e}_i) = 0$.

On the other hand, for any $x \in A$, the sequence

$$\frac{S(A_n - x, N(A_n) - \boldsymbol{e}_i)}{S(A_n - x_1, N(A_n) - \boldsymbol{e}_i)}$$

is bounded by $\max_j m_j(x_1)/m_j(x)$. In fact, if $\boldsymbol{r} \in R(n-1)$ and $B_1, \ldots, B_\ell$ are arbitrary disjoint subsets of $A_n - x_1$, with $\#B_i = r_i$ for all $i = 1, \ldots, \ell$, it is

$$\begin{aligned} S(A_n - x, \boldsymbol{r}) &= \sum_{B_1,\ldots,B_\ell} \prod_{x \in B_1} m_1(x) \cdots \prod_{x \in B_\ell} m_\ell(x)\, \frac{m_i(x_1)}{m_i(x)} \\ &\leq \max_j \frac{m_j(x_1)}{m_j(x)}\, S(A_n - x_1, \boldsymbol{r}) \end{aligned}$$

where, in each term, $i \in L$ is the index such that $x \in B_i$.

Therefore, if $\eta_i(x_1) = 0$ then we also have

$$\eta_i(x) = m_i(x)\, \lim_n \left[\frac{S(A_n - x, N(A_n) - \boldsymbol{e}_i)}{S(A_n - x_1, N(A_n) - \boldsymbol{e}_i)} \cdot \frac{S(A_n - x_1, N(A_n) - \boldsymbol{e}_i)}{S(A_n, N(A_n))} \right] = 0$$

for all $x \in A$. Assume now that for some $\omega \in \bar{\Omega}$, it is $\eta_i(x_1) = 0$ and $N_i(A) > 0$. Then

$$\eta_i(x) = \mathbb{P}\{N(x) = \boldsymbol{e}_i \mid \mathcal{G}_A\} = 0 \;\; \forall\, x \in A, \text{ and so } \mathbb{P}\{N_i(A) > 0 \mid \mathcal{G}_A\} = 0$$

(out of the negligible set where the latter probability does not match the sum of the former ones). But $\{N_i(A) > 0\} \in \mathcal{G}_A$, so that $\mathbb{1}_{\{N_i(A)>0\}}$ is a version of the latter conditional probability which takes the value 1 at ω.

Hence,

$$\Gamma_i = \bar{\Omega} \cap \{\eta_i(x_1) = 0,\; N_i(A) > 0\}$$

is a set of probability 0. □

Exclude from $\bar{\Omega}$ the negligible set $\cup_{i \in L} \Gamma_i$, so that $\bar{\Omega}$ will denote the resulting set of probability one in which (3.4) holds for all $i \in L$. For each nonempty $M \subset L$, let

$$\begin{aligned} \Omega_M &= \bar{\Omega} \cap \{N_i(A) > 0, \;\; \forall\, i \in M\} \cap \{N_i(A) = 0, \;\; \forall\, i \notin M\} \\ &= \bar{\Omega} \cap \{\xi_i(x_1) > 0, \;\; \forall\, i \in M\} \cap \{\xi_i(x_1) = 0, \;\; \forall\, i \notin M\}. \end{aligned}$$

On the one hand, Ω_M gives the set of trajectories of the random assignment process attributing indices exclusively in M to the elements of A. On the other hand, each Ω_M is in correspondence with the face Δ_M of Δ_ℓ defined as

$$\Delta_M = \{\boldsymbol{\xi} \in \Delta_\ell \;:\; \xi_i > 0, \;\; \forall\, i \in M \;\;\text{ and }\;\; \xi_i = 0, \;\; \forall\, i \notin M\}.$$

Remark 3 Obviously, we have $\bar{\Omega} = \bigcup_{M \subset L} \Omega_M$, but if $\mathbb{P}$ gives positive probability to several of them, say, $\Omega_{M_1}, \Omega_{M_2}, \ldots$, then $\mathbb{P}$ is the corresponding convex linear combination of

$$\mathbb{P}(\,\cdot \mid \Omega_{M_1}),\ \mathbb{P}(\,\cdot \mid \Omega_{M_2}), \ldots$$

Thus, when needed, we may suppose that $\mathbb{P}$ is concentrated on a single Ω_M and $\bar{\Omega} = \Omega_M$.

3.1 First asymptotic properties

Our next result uses the notations $m_{j|i}(x)$ and $m_{j|i}(B)$ introduced in Sect. 2. We note that a version of Lemma 3 was already used in Vélez and Prieto-Rumeau (2008). We also note that Lemma 3 is an algebraic result (meaning that it depends only on the algebraic properties of the weight function). It uses the notation $\boldsymbol{r} \in R(n)$ to denote a sequence $r(n)$, for $n \geq 1$, but, to simplify the notation, we will omit the dependence on n. The same convention will be applied in the forthcoming.

Lemma 3 *Let $A = \{x_n\}_{n \geq 1}$ be a sequence in E. Suppose that, for some indices $i, j \in L$, the sequence $m_{j|i}(x_n)$ has finite limit $m_{j|i} \in [0, \infty)$ as $n \to \infty$. Assume also that $\boldsymbol{r} \in R(n)$ varies with n in such a way that $r_i \to \infty$ as $n \to \infty$ and $r_j \geq 1$. Under these conditions,*

$$\lim_{n \to \infty} \frac{r_j\, S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_i)}{r_i\, S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_j)} = m_{j|i} \quad \forall\, x \in A. \tag{3.5}$$

Therefore, if $m_{j|i} > 0$, and in addition, $r_i/\varphi(n) \to y_i > 0$ and $r_j/\varphi(n) \to y_j > 0$ for some sequence $\varphi(n) \uparrow \infty$, then we have

$$\lim_{n \to \infty} \frac{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_i)}{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_j)} = \frac{y_i}{y_j}\, m_{j|i} \quad \forall\, x \in A. \tag{3.6}$$

Proof When n is large enough so that $x \in A_n$, the following identity is readily established:

$$\sum_{x_q \in A_n - x} m_j(x_q)\, S(A_n - x - x_q, \boldsymbol{r} - \boldsymbol{e}_i - \boldsymbol{e}_j) = r_j\, S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_i)$$

(to see this, just count the number of terms in each side). Thus

$$\frac{r_j\, S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_i)}{r_i\, S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_j)} = \sum_{x_q \in A_n - x}^{n} m_{j|i}(x_q)\, a_n(x_q),$$

with

$$a_n(x_q) = \frac{m_i(x_q) S(A_n - x - x_q, \boldsymbol{r} - \boldsymbol{e}_i - \boldsymbol{e}_j)}{r_i\, S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_j)}.$$

We note that the $a_n(x_q)$ satisfy

$$\sum_{x_q \in A_n - x} a_n(x_q) = 1 \quad \text{and} \quad \lim_{n\to\infty} a_n(x_q) \to 0$$

for every fixed q provided that $r_i \to \infty$. Indeed,

$$S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_j) = \sum_{i=0}^{\ell} m_i(x_q) S(A_n - x_q, \boldsymbol{r} - \boldsymbol{e}_j - \boldsymbol{e}_i)$$

yields $r_i a_n(x_q) \leq 1$. Hence, (3.5) is a consequence of the Toeplitz lemma (see, e.g., Loève 1977, p. 250), while (3.6) easily follows. □

Proposition 2 *Let* $\boldsymbol{Y}$ *be a random assignment process on* $(E, \boldsymbol{m})$ *and let* A *be any sequence in* E.

(a) If $N_i(A) = \infty$ *and* $m_{j|i}(x_n)$ *has a finite limit* $m_{j|i} \in [0, \infty)$ *as* $n \to \infty$ *then, almost surely,*

$$\lim_{n\to\infty} \frac{N_j(A_n)}{N_i(A_n)} = \frac{\xi_j(x)}{\xi_i(x)}\, m_{j|i} \quad \forall\, x \in A. \tag{3.7}$$

Hence, $\xi_j(x)/\xi_i(x)$ *is a.s. constant (that is, it does not depend on* $x \in A$*) when* $m_{j|i} > 0$.

(b) If $\limsup_{n\to\infty} N_i(A_n)/n > 0$ *and* $\lim_{n\to\infty} m_{j|i}(x_n) = \infty$, *then* $N_j(A) < \infty$ *almost surely.*

Proof (a) Assume that $\omega \in \bar{\Omega}$ is such that $N_i(A) = \infty$, and therefore, $\xi_i(x) > 0$ for all $x \in A$ (see Proposition 1). If $N_j(A) = 0$ and $\xi_j(x) = 0$, then the stated result is obvious. Otherwise, suppose that $\xi_j(x) > 0$ and $N_j(A) > 0$. Applying Lemma 3 with $\boldsymbol{r} = N(A_n)$, we get

$$\frac{N_j(A_n)}{N_i(A_n)} \, \frac{S(A_n - x, N(A_n) - \boldsymbol{e}_i)}{S(A_n - x, N(A_n) - \boldsymbol{e}_j)} \longrightarrow m_{j|i},$$

and as a consequence of (3.3), the limit (3.7) indeed holds.

(b) To prove this statement, we proceed by contradiction. So, we suppose that $m_{j|i}(x_n) \to \infty$ and $N_j(A) = \infty$. We apply the case (a) above by interchanging the roles of i and j so as to obtain

$$\frac{N_i(A_n)}{N_j(A_n)} \cdot \frac{S(A_n - x, N(A_n) - \boldsymbol{e}_j)}{S(A_n - x, N(A_n) - \boldsymbol{e}_i)} \longrightarrow 0,$$

which yields

$$\frac{N_i(A_n)}{N_j(A_n)} \longrightarrow 0.$$

Now, since $N_j(A_n) \leq n$, the above limit is not compatible with the fact that $\limsup_{n\to\infty} N_i(A_n)/n$ is positive. We conclude that, necessarily, $N_j(A)$ is finite. $\square$

Remark 4 Under condition (A)—recall Sect. 2—the limit

$$\boldsymbol{m} = \lim_{n\to\infty} \boldsymbol{m}(x_n) \in (0,\infty)^L$$

exists and $m_{j|i} = m_j/m_i > 0$ for all $i, j \in L$. We thus conclude from Proposition 2(a) that $\xi_i(x) = \xi_i$ for every $x \in A$, where $\boldsymbol{\xi} \in \Delta_\ell$ is $\mathcal{G}_A$-measurable. Moreover, $N_j(A_n)/N_i(A_n) \to \xi_j m_j/\xi_i m_i$ and since

$$\big(N_0(A_n)/n, N_1(A_n)/n, \ldots, N_\ell(A_n)/n\big)$$

must have some limit point in Δ_ℓ, we easily get that $N_i(A_n)/n \to \xi_i m_i/\boldsymbol{\xi}\cdot\boldsymbol{m}$ almost surely for all $i \in L$. This is the strong law of large numbers formulated in Theorem 3.4 in Vélez and Prieto-Rumeau (2008).

Our next lemma is, in fact, an algebraic result (which does not depend on the random assignment process itself). The notation, however, is greatly simplified using conditional expectations related to the random assignment process. Lemma 4 below allows to derive a first bound on the growth rate of $N_j(A_n)$, which is given in Proposition 3.

Lemma 4 *Let A be any sequence in E. For $\boldsymbol{r} \in R(n)$ such that $r_i, r_j \geq 1$ and every $x \in A$ we have*

$$r_j\,\frac{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_i)}{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_j)} = \mathbb{E}[m_{j|i}(A_n^i) \mid N(A_n - x) = \boldsymbol{r} - \boldsymbol{e}_j] \tag{3.8}$$

where A_n^i is the random set $\{x' \in A_n - x : N_i(x') = 1\}$.

Proof Let $x_0 \in A_n$ be fixed. If $B_1, \ldots, B_\ell$ are arbitrary disjoint subsets of $A_n - x_0$ with $\#B_k = r_k$, for $k \neq j$, and $\#B_j = r_j - 1$, the following identity follows:

$$\sum_{B_1,\ldots,B_\ell} \prod_{x\in B_1} m_1(x)\cdots \prod_{x\in B_\ell} m_\ell(x)\, m_{j|i}(B_i) = r_j\, S(A_n - x_0, \boldsymbol{r} - \boldsymbol{e}_i). \tag{3.9}$$

Indeed, the sum in the lefthand side would give $S(A_n - x_0, \boldsymbol{r} - \boldsymbol{e}_j)$ in the absence of $m_{j|i}(B_i)$. This latter factor replaces $m_i(x)$ with $m_j(x)$ for some $x \in B_i$, augmenting B_j with x and reducing B_i accordingly, so that each term of $S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_i)$ is repeated r_j times.

Now, according to (2.2),

$$\prod_{x\in B_1} m_1(x)\cdots \prod_{x\in B_\ell} m_\ell(x)/S(A_n - x_0, \boldsymbol{r} - \boldsymbol{e}_j)$$

is the probability, conditional on $N(A_n - x_0) = \boldsymbol{r} - \boldsymbol{e}_j$, that the indices $1, \ldots, \ell$ will be attributed to the points in $B_1, \ldots, B_\ell$, respectively. This completes the proof. $\square$

Proposition 3 *Let $\boldsymbol{Y}$ be a random assignment process on $(E, \boldsymbol{m})$ and let A be any sequence in E. For all $i, j \in L$ such that $N_i(A) > 0$ and $N_j(A) > 0$, we have almost surely*

$$\limsup_{n\to\infty} \frac{N_j(A_n)}{m_{j|i}(A_n)} < \frac{\xi_j(x)}{\xi_i(x)} \qquad \forall\, x \in A.$$

Proof Since the righthand term of (3.8) is bounded by $m_{j|i}(A_n)$, by replacing $\boldsymbol{r}$ with $N(A_n)$ we get

$$N_j(A_n)\,\frac{S(A_n - x, N(A_n) - \boldsymbol{e}_i)}{S(A_n - x, N(A_n) - \boldsymbol{e}_j)} \le m_{j|i}(A_n).$$

Now, according to (3.3), the stated result follows. □

The relevant conclusion of this proposition is that $N_j(A_n)/m_{j|i}(A_n)$ remains bounded. We make use of this fact in the next result.

Corollary 2 *Let $\boldsymbol{Y}$ be a random assignment process on $(E, \boldsymbol{m})$ and let A be any sequence in E.*

(a) If $i, j \in L$ are such that $N_i(A) > 0$ and $m_{j|i}(A) < \infty$, then $N_j(A) < \infty$ almost surely.
(b) If $i, j \in L$ are such that $\limsup_{n\to\infty} N_i(A_n)/n > 0$ and $\lim_{n\to\infty} m_{i|j}(x_n) = 0$, then $N_j(A) = 0$ almost surely.

Proof The statement (a) is obvious. Concerning (b), assume that $N_j(A) > 0$. Then we would have, almost surely,

$$\limsup_{n\to\infty} \frac{N_i(A_n)}{m_{i|j}(A_n)} < \infty.$$

But, on the other hand,

$$\frac{m_{i|j}(A_n)}{n} \sim m_{i|j}(x_n) \to 0,$$

which contradicts the fact that $\limsup_{n\to\infty} N_i(A_n)/n > 0$. □

3.2 Regular sequences

Next, we introduce the concept of a regular sequence. We say that a sequence $A = \{x_n\}_{n\ge 1}$ in E is regular if the limit

$$\lim_{n\to\infty} m_{j|i}(x_n) = m_{j|i}$$

exists (with $0 \le m_{j|i} \le \infty$) for all $i, j \in L$. Then we can define the following equivalence relation on L:

$$i \sim j \iff m_{j|i} \in (0, \infty).$$

Furthermore, the equivalence classes may be totally ordered by means of the strict order relation:

$$j \prec i \iff m_{j|i} = 0.$$

Summing up, the set L is totally pre-ordered by the growth rate of the sequences $m_{j|i}(A_n)$. This order, of course, depends on the regular sequence A.

Let A be a regular sequence. As explained in Remark 3, we will now suppose that the probability measure $\mathbb{P}$ is such that $\mathbb{P}(\Omega_M) = 1$ for some fixed $M \subset L$. We may thus assume that $N_i(A) > 0$ and $\xi_i(x) > 0$ for all $i \in M$ and $x \in A$, whereas $N_i(A) = 0$ and $\xi_i(x) = 0$ if $i \notin M$ and $x \in A$.

The set M can be decomposed as

$$M = C_0 \cup C_1 \cup \cdots \cup C_s \tag{3.10}$$

where $C_0, C_1, \ldots, C_s$ are ordered equivalence classes of the indices in M. That means $i \sim j$ when $i, j \in C_p$, and $j \prec i$ if $i \in C_p$ and $j \in C_{p+1}$. We will also use the notation $j \prec C_p$, meaning that $j \prec i$ for all $i \in C_p$ (or equivalently, $j \in C_k$ for some $k > p$). Similarly, we will write $C_0 \succ C_1 \succ \cdots \succ C_s$.

It is obvious that a random assignment process remains unchanged when each $m_i(x)$ is replaced with $c_i \cdot m_i(x)$ for some constant $c_i \in (0, \infty)$. Thus, to avoid a more cumbersome notation, let us choose an arbitrary index i in each C_p and replace $m_j(x)$ with $m_{i|j} m_j(x)$ for all $j \in C_p$. Then we have $m_{i|j} = 1$ whenever $i \sim j$.

4 Proof of the main results

Our next results are aimed towards establishing a strong law of large numbers.

Let A be a regular sequence and suppose that the probability measure $\mathbb{P}$ is such that $\mathbb{P}(\Omega_M) = 1$ for some fixed $M \subset L$. Clearly, there is some $i \in M$ with $\limsup_{n\to\infty} N_i(A_n)/n > 0$. Then $i \in C_0$ [(recall (3.10)] since, according to part (b) of Corollary 2, $i \in M - C_0$ would imply $N_j(A) = 0$ for all $j \in C_0$. Moreover, by Proposition 2(a), we have, almost surely as $n \to \infty$,

$$\frac{N_{i'}(A_n)}{N_i(A_n)} \to \nu_{i'|i} \in (0, \infty) \quad \forall\, i' \in C_0 \qquad \text{and} \qquad \frac{N_j(A_n)}{N_i(A_n)} \to 0 \quad \forall\, j \in M - C_0.$$

Now, $\sum_{j\in M} N_j(A_n) = n$ may be written as

$$1 = \frac{N_i(A_n)}{n} \sum_{i'\in C_0} \frac{N_{i'}(A_n)}{N_i(A_n)} + \frac{N_i(A_n)}{n} \sum_{j\in M-C_0} \frac{N_j(A_n)}{N_i(A_n)},$$

and taking into account the preceding remarks, this yields (almost surely)

$$\lim_{n\to\infty} \frac{N_i(A_n)}{n} = \nu_i > 0 \quad \forall\, i \in C_0 \qquad \text{with} \qquad \sum_{i\in C_0} \nu_i = 1.$$

Comparing with (3.7), for any $i, j \in C_0$ we have $\xi_j(x)/\xi_i(x) = \nu_j/\nu_i$. Therefore, for every $x \in A$,

$$\xi_i(x) = \alpha_0(x)\, \nu_i \qquad \text{where} \qquad \alpha_0(x) = \sum_{i\in C_0} \xi_i(x).$$

Call $\bar{N}_0(A_n) = \sum_{i\in C_0} N_i(A_n)$. Then we have $\bar{N}_0(A_n)/n \to 1$, but it may happen that $n - \bar{N}_0(A_n)$ remains bounded (for instance, if $m_{j|i}(A) < \infty$ for all $i \in C_0$ and $j \in M - C_0$, it will be $N_j(A) < \infty$ for all $j \in M - C_0$). On the contrary, when $m_{j|i}(A) = \infty$ for some $j \in C_1$ and any $i \in C_0$, we may have $N_j(A) = \infty$. In this case, the assertion of Proposition 2(a) gives, almost surely as $n \to \infty$,

$$\frac{N_{j'}(A_n)}{N_j(A_n)} \to \nu_{j'|j} \in (0, \infty) \quad \forall\, j' \in C_1 \qquad \text{and} \qquad \frac{N_k(A_n)}{N_j(A_n)} \to 0 \quad \forall\, k \prec C_1.$$

Then, the equality

$$1 = \frac{N_j(A_n)}{n - \bar{N}_0(A_n)} \sum_{j'\in C_1} \frac{N_{j'}(A_n)}{N_j(A_n)} + \frac{N_j(A_n)}{n - \bar{N}_0(A_n)} \sum_{k\prec C_1} \frac{N_k(A_n)}{N_j(A_n)}$$

shows that almost surely

$$\lim_{n\to\infty} \frac{N_j(A_n)}{n - \bar{N}_0(A_n)} = \nu_j > 0 \quad \forall\, j \in C_1 \qquad \text{with} \qquad \sum_{j\in C_1} \nu_j = 1.$$

Again, the comparison with (3.7) yields, for all $j \in C_1$ and $x \in A$,

$$\xi_j(x) = \alpha_1(x)\nu_j \qquad \text{where} \qquad \alpha_1(x) = \sum_{j\in C_1} \xi_j(x).$$

Letting $\bar{N}_p(A_n) = \sum_{j\in C_p} N_j(A_n)$, the argument above may be pursued similarly. At each step it may happen that the indices j of the next C_{p+1} are such that $N_j(A) < \infty$ (a sufficient condition for this is that $m_{j|i}(A) < \infty$ for some $i \succ j$). Or, if $m_{j|i}(A) = \infty$ for all $i \succ j$, it may be $N_j(A) = \infty$, and in this case, almost surely as $n \to \infty$,

$$\frac{N_j(A_n)}{n - \sum_{t\le p} \bar{N}_t(A_n)} \to \nu_j > 0 \quad \forall\, j \in C_{p+1} \qquad \text{with} \qquad \sum_{j\in C_{p+1}} \nu_j = 1.$$

Furthermore, we have, for all $j \in C_{p+1}$ and $x \in A$,

$$\xi_j(x) = \alpha_{p+1}(x)\, \nu_j,$$

where $\alpha_{p+1}(x) = \sum_{j \in C_{p+1}} \xi_j(x)$.

We will assume that $M \subset L$ is a set of indices, which is decomposed into ordered equivalence classes; recall (3.10). In what follows, we will suppose that the regular sequence A satisfies, in addition, the condition (A*). This allows to have $N_i(A) = \infty$ for all $i \in M$, and we will show that this indeed holds.

Lemma 5 *Assume that the sequence A verifies (A*) and that $\boldsymbol{r} \in R(n)$ varies with n in such a way that, for all $i \in C_0$, $j \in C_1$, and $k \prec C_1$, as $n \to \infty$ it is*

$$r_i/n \to y_i, \quad r_j/m_{j|i}(A_n) \to y_j, \quad \textit{and} \quad r_k/m_{j|i}(A_n) \to 0,$$

where $y_i > 0$ for every $i \in C_0$ and $\sum_{i \in C_0} y_i = 1$. Then, for any $i \in C_0$ and $j \in C_1$, we have

$$\lim_{n \to \infty} \frac{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_j)}{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_i)} = \frac{y_j}{y_i} \quad \forall\, x \in A$$

provided that the latter limit is finite for some $i \in C_0$.

Proof Let $i \in C_0$ be such that

$$\lim_{n \to \infty} \frac{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_j)}{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_i)} = \beta$$

for some finite β. Summing up the equalities (3.8) for the different values of $i' \in C_0$ we get

$$\frac{r_j}{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_j)} \sum_{i' \in C_0} S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_{i'}) = \mathbb{E}\Big[\sum_{i' \in C_0} m_{j|i'}(A_n^{i'}) \Big|\, D_n \Big], \tag{4.1}$$

where $D_n = \{N(A_n - x) = \boldsymbol{r} - \boldsymbol{e}_j\}$. According to (3.6) and recalling that $m_{i'|i} = 1$, the lefthand term of (4.1) is of order $y_j m_{j|i}(A_n)/(\beta\, y_i)$ and diverges when $n \to \infty$ for $j \in C_1$. Regarding the righthand term, fix any $\varepsilon > 0$ and let q_ε be such that $1 - \varepsilon < m_{i|i'}(x_q) < 1 + \varepsilon$ for all $q \geq q_\varepsilon$. Then we have

$$(1 - \varepsilon) m_{j|i}(A_n^{i'} - A_{q_\varepsilon}) \leq m_{j|i'}(A_n^{i'} - A_{q_\varepsilon}) \leq (1 + \varepsilon) m_{j|i}(A_n^{i'} - A_{q_\varepsilon})$$

and

$$(1 - \varepsilon) m_{j|i}\big(\bar{A}_n^0 - A_{q_\varepsilon}\big) + T \leq \sum_{i' \in C_0} m_{j|i'}(A_n^{i'}) \leq (1 + \varepsilon) m_{j|i}\big(\bar{A}_n^0 - A_{q_\varepsilon}\big) + T,$$

where T is the bounded term $\sum_{i'\in C_0} m_{j|i'}(A_n^{i'}\cap A_{q_\varepsilon})$ and $\bar{A}_n^0=\cup_{i'\in C_0}A_n^{i'}$. Therefore,

$$\begin{aligned}\mathbb{E}[(1-\varepsilon)m_{j|i}\big(\bar{A}_n^0-A_{q_\varepsilon}\big)+T\mid D_n]&\le \mathbb{E}\Bigg[\sum_{i'\in C_0} m_{j|i'}(A_n^{i'})\big|D_n\Bigg]\\&\le \mathbb{E}\Bigg[(1+\varepsilon)m_{j|i}\big(\bar{A}_n^0-A_{q_\varepsilon}\big)+T\mid D_n\Bigg],\end{aligned}$$

and the divergence of the lefthand term of (4.1) shows that $\mathbb{E}[m_{j|i}(\bar{A}_n^0)\mid D_n]\to\infty$. Hence, if the above inequality is divided by $\mathbb{E}[m_{j|i}(\bar{A}_n^0)\mid D_n]$, we get

$$\mathbb{E}\Bigg[\sum_{i'\in C_0} m_{j|i'}(A_n^{i'})\mid D_n\Bigg]\sim \mathbb{E}[m_{j|i}(\bar{A}_n^0)\mid D_n].$$

Now, if B_n is an arbitrary subset of A_n-x with

$$\#B_n=\sum_{j\prec C_0} r_j,$$

it is

$$m_{j|i}(A_n)-\max_{B_n} m_{j|i}(B_n)\le \mathbb{E}[m_{j|i}(\bar{A}_n^0)\mid D_n]\le m_{j|i}(A_n)-\min_{B_n} m_{j|i}(B_n).$$

Moreover, if $m_{j|i}(x_q)<\varepsilon$ for $q\ge q_\varepsilon$, we have

$$m_{j|i}(B_n)\le m_{j|i}(A_{q_\varepsilon})+\varepsilon\sum_{k\prec C_0} r_k \qquad\text{and}\qquad \frac{\max_{B_n} m_{j|i}(B_n)}{m_{j|i}(A_n)}\to 0.$$

Consequently, $\mathbb{E}[m_{j|i}(\bar{A}_n^0)\mid D_n]\sim m_{j|i}(A_n)$. Finally, (4.1) gives $\beta=y_j/y_i$, and more generally, the stated result holds for all $i\in C_0$. $\square$

Our first results towards the law of large numbers are Proposition 4 and its continuation, Proposition 5.

Proposition 4 *Let $\boldsymbol{Y}$ be a random assignment process on $(E,\boldsymbol{m})$ and let A be a sequence in E that verifies the condition ($A^\star$). Assume further that $\mathbb{P}(\Omega_M)=1$ for some $M\subset L$ (recall Remark 3), and that $M=\cup_{p=0}^s C_p$, where $C_0\succ C_1\succ\cdots\succ C_s$ are equivalence classes of indices [(see (3.10)]. Then, with probability 1, for all $x\in A$,*

(a) $\displaystyle\lim_{n\to\infty}\frac{N_i(A_n)}{n}=\nu_i=\frac{\xi_i(x)}{\alpha_0(x)}$ *for all $i\in C_0$.*

(b) $\displaystyle\lim_{n\to\infty}\frac{N_j(A_n)}{m_{j|i}(A_n)}=\frac{\alpha_1(x)\nu_j}{\alpha_0(x)}=c_1\nu_j$ *for all $j\in C_1$ and $i\in C_0$.*

Proof Part (a) has already been proved. Now, Proposition 3 ensures that there are subsequences of n with $N_j(A_n)/m_{j|i}(A_n) \to y_j$ for all $j \in C_1$ and any $i \in C_0$. According to Prosposition 2(a), along these subsequences, $N_k(A_n)/m_{j|i}(A_n) \to 0$ for all $k \prec C_1$ and

$$\frac{S(A_n - x, N(A_n) - \boldsymbol{e}_j)}{S(A_n - x, N(A_n) - \boldsymbol{e}_i)} \longrightarrow \frac{\xi_j(x)}{\xi_i(x)} = \frac{\alpha_1(x)\nu_j}{\alpha_0(x)\nu_i} \quad \forall\, x \in A$$

when $j \in C_1$ and $i \in C_0$. Lemma 5 is therefore applicable to the corresponding subsequence of $\boldsymbol{r} = N(A_n)$ and gives

$$\frac{S(A_n - x, N(A_n) - \boldsymbol{e}_j)}{S(A_n - x, N(A_n) - \boldsymbol{e}_i)} \longrightarrow \frac{y_j}{\nu_i} \quad \forall\, x \in A.$$

Hence, y_j equals $\alpha_1(x)\nu_j/\alpha_0(x)$, and the whole sequence converges:

$$\frac{N_j(A_n)}{m_{j|i}(A_n)} \longrightarrow \frac{\alpha_1(x)\nu_j}{\alpha_0(x)} \quad \forall\, x \in A.$$

It then follows that $\alpha_1(x)/\alpha_0(x)$ must have a constant value c_1, which does not depend on $x \in A$. □

Lemma 6 *Assume that the sequence A verifies the condition ($A^\star$) and that $\boldsymbol{r} \in R(n)$ varies with n in such a way that, for all $i \in C_0$, $j \in C_1$, $k \in C_2$, and $l \prec C_2$, and as $n \to \infty$,*

$$r_i/n \to y_i,\; r_j/m_{j|i}(A_n) \to y_j,\; r_k/m_{k|i}(A_n) \to y_k,\; \textit{and}\;\; r_l/m_{k|i}(A_n) \to 0,$$

where $y_i > 0$ for each $i \in C_0 \cup C_1$ and $\sum_{i \in C_0} y_i = 1$. Then, for any $i \in C_0$ and $k \in C_2$,

$$\lim_{n\to\infty} \frac{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_k)}{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_i)} = \frac{y_k}{y_i} \quad \forall\, x \in A,$$

provided that the limit is finite for some $i \in C_0$.

Proof As in Lemma 5 we obtain, for any $k \in C_2$ and $i \in C_0$,

$$\frac{r_k}{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_k)} \sum_{i \in C_0} S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_i) \sim \mathbb{E}[m_{k|i}(\bar{A}_n^0) \mid D_n], \tag{4.2}$$

where now we let $D_n = \{N(A_n - x) = \boldsymbol{r} - \boldsymbol{e}_k\}$. But

$$\begin{aligned} m_{k|i}(A_n) &\geq \mathbb{E}[m_{k|i}(\bar{A}_n^0) \mid D_n] \\ &\geq m_{k|i}(A_n) - \sum_{j \in C_1} \mathbb{E}[m_{k|i}(A_n^j) \mid D_n] - \sum_{k \prec C_1} \mathbb{E}[m_{k|i}(A_n^k) \mid D_n], \end{aligned}$$

and for the last terms, if $m_{k|i}(x_q) < \varepsilon$ when $q \geq q_\varepsilon$, it is $\mathbb{E}[m_{k|i}(A_n^k) \mid D_n] \leq m_{k|i}(A_{q_\varepsilon}) + \varepsilon r_k$, so that $\mathbb{E}[m_{k|i}(A_n^k) \mid D_n]/m_{k|i}(A_n) \to 0$. For the terms corresponding to $j \in C_1$, if $m_{j|i}(x_q) < \varepsilon$ as well for $q \geq q_\varepsilon$, it is

$$\mathbb{E}[m_{k|i}(A_n^j) \mid D_n] \leq m_{k|i}(A_{q_\varepsilon}) + \varepsilon \mathbb{E}[m_{k|j}(A_n^j) \mid D_n].$$

Now, from (3.8) applied to (k, i) and (k, j) and Lemma 5, we get

$$\frac{\mathbb{E}[m_{k|j}(A_n^j) \mid D_n]}{\mathbb{E}[m_{k|i}(A_n^i) \mid D_n]} = \frac{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_j)}{S(A_n - x, \boldsymbol{r} - \boldsymbol{e}_i)} \longrightarrow \frac{y_j}{y_i}$$

and $\mathbb{E}[m_{k|i}(A_n^j) \mid D_n]/m_{k|i}(A_n) \to 0$. Therefore,

$$\mathbb{E}[m_{k|i}(\bar{A}_n^0) \mid D_n] \sim m_{k|i}(A_n).$$

Together with (4.2), this easily gives the stated result. □

Our next result is the second part of Proposition 4.

Proposition 5 *Under the conditions of Proposition 4, with probability* 1 *and for all* $x \in A$,

(c) $\displaystyle\lim_{n\to\infty} \frac{N_k(A_n)}{m_{k|i}(A_n)} = \frac{\alpha_2(x)\nu_k}{\alpha_0(x)} = c_2\nu_k$ *for all* $k \in C_2$ *and* $i \in C_0$.

(d) *For* $p > 2$, $\displaystyle\lim_{n\to\infty} \frac{N_l(A_n)}{m_{l|i}(A_n)} = \frac{\alpha_p(x)\nu_l}{\alpha_0(x)} = c_p\nu_l$ *for all* $l \in C_p$ *and* $i \in C_0$.

As a consequence, $\alpha_p(x)$ *has a constant value* α_p *for each* $p = 1, \ldots, s$, *and* $\xi_i(x)$ *has also a constant value* ξ_i *for each* $i \in M$.

Proof From Proposition 4,

$$N_i(A_n)/n \to \nu_i \quad \forall\, i \in C_0 \qquad \text{and} \qquad N_j(A_n)/m_{j|i}(A_n) \to c_1\nu_j \quad \forall\, j \in C_1.$$

As a consequence of Proposition 3, there are subsequences of n with

$$N_k(A_n)/m_{k|i}(A_n) \to y_k$$

for each $k \in C_2$ and $i \in C_0$; and along such subsequences $N_l(A_n)/m_{k|i}(A_n) \to 0$ for each $l \prec C_2$. Furthermore, for $x \in A$,

$$\frac{S(A_n - x, N(A_n) - \boldsymbol{e}_k)}{S(A_n - x, N(A_n) - \boldsymbol{e}_i)} \longrightarrow \frac{\xi_k(x)}{\xi_i(x)} = \frac{\alpha_2(x)\nu_k}{\alpha_0(x)\nu_i}$$

when $k \in C_2$ and $i \in C_0$. Lemma 6 can be therefore applied to the corresponding subsequence of $\boldsymbol{r} = N(A_n)$ and gives

$$\frac{S(A_n - x, N(A_n) - \boldsymbol{e}_k)}{S(A_n - x, N(A_n) - \boldsymbol{e}_i)} \longrightarrow \frac{y_k}{\nu_i}.$$

Then the whole sequence converges: $N_k(A_n)/m_{k|i}(A_n) \to \alpha_2(x)\nu_k/\alpha_0(x)$ for all $x \in A$. Hence, $\alpha_2(x)/\alpha_0(x)$ must have a constant value c_2.

The proof of (d) requires a generalization of Lemma 6, using similar arguments, and it is omitted.

Finally, since $1 = \sum_{p=0}^{s} \alpha_p(x) = \alpha_0(x)\left(1 + \sum_{p=1}^{s} c_p\right)$, we conclude that $\alpha_0(x) = \alpha_0$ and $\alpha_p(x) = \alpha_0 c_p = \alpha_p$. Hence, for all $i \in M$, $\xi_i(x) = \alpha_p \nu_i = \xi_i$. $\square$

The previous results may be summarized in the following strong law of large numbers.

Theorem 2 *If $\boldsymbol{Y}$ is a random assignment process on $(E, \boldsymbol{m})$ and A is a sequence that verifies the condition ($A^\star$), then there exists a $\mathcal{G}_A$-measurable probability vector $\boldsymbol{\xi} \in \Delta_\ell$ such that given a subset $M \subset L$ with $\mathbb{P}(\Omega_M) > 0$, and such that $M = \cup_{p=0}^{s} C_p$, where $C_0 \succ C_1 \succ \cdots \succ C_s$ are the corresponding ordered equivalence classes, we have, almost surely in Ω_M,*

$$\lim_{n\to\infty} \frac{N_i(A_n)}{n} = \frac{\xi_i}{\alpha_0} \quad \forall\, i \in C_0$$

and

$$\lim_{n\to\infty} \frac{N_j(A_n)}{m_{j|i}(A_n)} = \frac{\xi_j}{\alpha_0} \quad \forall\, i \in C_0,\ j \in M - C_0,$$

where $\alpha_0 = \sum_{i \in C_0} \xi_i$, while $N_j(A) = \xi_j = 0$ when $j \notin M$.

Next, we will establish the De Finetti theorem. Our next lemma is a consequence of Lemmas 5 and 6.

Lemma 7 *Assume that the sequence A verifies the condition ($A^\star$). Let $M \subset L$ be a subset of indices [(recall (3.10)], and suppose that $\boldsymbol{r} \in R(n)$ varies with n in such a way that, as $n \to \infty$,*

$$r_i/n \to y_i \quad \forall\, i \in C_0 \qquad \text{and} \qquad r_j/m_{j|i}(A_n) \to y_j \quad \forall\, j \in M - C_0,$$

where $y_i > 0$ for every $i \in M$ and $\sum_{i \in C_0} y_i = 1$.

(a) For all $i, j \in M$,

$$\lim_{n\to\infty} \frac{S(A_n, \boldsymbol{r} - \boldsymbol{e}_j)}{S(A_n, \boldsymbol{r} - \boldsymbol{e}_i)} = \frac{y_j}{y_i}.$$

(b) Let $y_k = 0$ for the indices $k \notin M$. Then, for all $B \in \mathcal{B}$ with $B \subset E - A$, and any $\boldsymbol{k} \in R(\#B)$ such that $k_i = 0$ for $i \notin M$, it is

$$\lim_{n\to\infty} \frac{S(A_n, \boldsymbol{r} - \boldsymbol{k})}{S(B \cup A_n, \boldsymbol{r})} = \frac{\prod_{i \in M} y_i^{k_i}}{\prod_{x \in B} \boldsymbol{y} \cdot \boldsymbol{m}(\boldsymbol{x})}. \tag{4.3}$$

where $k_0 = \#B - \bar{k}$ if $0 \in M$.

Proof (a) If $A = \{x_q\}_{q\geq 1}$ is a regular sequence satisfying (A*), then so is $A' = \{x'_q\}_{q\geq 1}$ given by $x'_1 = x$ for any $x \in E$, and $x'_q = x_{q-1}$ for $q \geq 2$. Then, Lemmas 5 and 6 applied to A' and x yield the stated result.

(b) We begin with the case $B = \{x\}$ and $\boldsymbol{k} = \boldsymbol{e}_j$ with $j \in M$. Then, by part (a),

$$\frac{S(A_n \cup \{x\}, \boldsymbol{r})}{S(A_n, \boldsymbol{r} - \boldsymbol{e}_j)} = \sum_{i\in M} m_i(x) \frac{S(A_n, \boldsymbol{r} - \boldsymbol{e}_i)}{S(A_n, \boldsymbol{r} - \boldsymbol{e}_j)} \longrightarrow \sum_{i\in M} m_i(x) \frac{y_i}{y_j} = \frac{\boldsymbol{y}\cdot \boldsymbol{m}(x)}{y_j}.$$

Now, if $B = \{x, x', \ldots\}$ then the limit (4.3) follows by recursively considering the sequences $A \cup \{x\}$, $A \cup \{x, x'\}$, etc. □

Now we are ready to prove the De Finetti theorem.

4.1 Proof of Theorem 1

First assume that $\mathbb{P}(\Omega_M) = 1$ for some subset $M \subset L$ (see Remark 3). For any $B \in \mathcal{B}$, the sequence A may be shifted to get $B \subset E - A$. Then, apply Lemma 2 with $C = B$ to get, for $\boldsymbol{k} \in R(\#B)$,

$$\mathbb{P}\{N(B) = \boldsymbol{k} \mid \mathcal{G}_{B\cup A}\} = S(B, \boldsymbol{k}) \lim_{n\to\infty} \frac{S(A_n, N(B\cup A_n) - \boldsymbol{k})}{S(B\cup A_n, N(B\cup A_n))}.$$

But, according to Theorem 2, for some $\mathcal{G}_{A\cup B}$-measurable (and so, $\mathcal{G}_A$-measurable) probability vector $\boldsymbol{\xi} \in \Delta_M$

$$N_i(B\cup A_n)/n \to \xi_i/\alpha_0 \quad \forall\, i \in C_0 \qquad \text{and} \qquad N_j(B\cup A_n)/m_{j|i}(A_n) \to \xi_j/\alpha_0$$

for all $j \in M - C_0$. Thus, Lemma 7 applied to the sequence $\boldsymbol{r} = N(B\cup A_n)$ gives

$$\mathbb{P}\{N(B) = \boldsymbol{k} \mid \mathcal{G}_{B\cup A}\} = S(B, \boldsymbol{k}) \frac{\prod_{i\in L} \xi_i^{k_i}}{\prod_{x\in B} \boldsymbol{\xi}\cdot \boldsymbol{m}(x)}.$$

(Note that if $k_i > 0$ for some $i \notin M$, then the above conditional probability vanishes because $\xi_i = 0$, which is consistent with the definition of Ω_M.) Since $\boldsymbol{\xi}$ is $\mathcal{G}_{A\cup B}$-measurable,

$$\mathbb{P}\{N(B) = \boldsymbol{k} \mid \boldsymbol{\xi}\} = \mathbb{E}[\mathbb{P}\{N(B) = \boldsymbol{k} \mid \mathcal{G}_{B\cup A}\} \mid \boldsymbol{\xi}] = S(B, \boldsymbol{k}) \frac{\prod_{i\in L} \xi_i^{k_i}}{\prod_{x\in B} \boldsymbol{\xi}\cdot \boldsymbol{m}(x)}. \tag{4.4}$$

Now, from (2.2), for any $B \in \mathcal{B}$, $\boldsymbol{r} \in R(\#B)$ with $r_i = 0$ for $i \notin M$, and any disjoint subsets B_i with $\#B_i = r_i$, for $i \in M$, we get

$$\mathbb{P}\left\{\bigcap_{i\in M}\{Y(x) = i,\ \forall x \in B_i\} \mid \boldsymbol{\xi}\right\} = \mathbb{P}\{N(B) = \boldsymbol{r} \mid \boldsymbol{\xi}\}\,\frac{1}{S(B,\boldsymbol{r})}\prod_{i\in M}\prod_{x\in B_i} m_i(x)$$
$$= \frac{\prod_{i\in M}\prod_{x\in B_i}\xi_i m_i(x)}{\prod_{x\in B}\boldsymbol{\xi}\cdot\boldsymbol{m}(x)},$$

which shows the conditional independence of $\{Y(x)\}_{x\in E}$ given $\boldsymbol{\xi}$.

In the case of a general probability measure $\mathbb{P}$ (not necessarily supported on some Ω_M), since $\mathbb{P}$ is the convex linear combination of several $\mathbb{P}_{M_1}, \mathbb{P}_{M_2}, \ldots$ concentrated on disjoint $\Omega_{M_1}, \Omega_{M_2}, \ldots$, one can define a unique $\boldsymbol{\xi} \in \Delta_\ell$ that coincides, in each Ω_M, with the corresponding $\boldsymbol{\xi}$. The conditional independence given $\boldsymbol{\xi}$ is then obtained by linearity, as well as (4.4). Now, (2.9) easily follows and the proof of Theorem 1 is now complete.

4.2 Proof of Corollary 1

Assume that $\mathcal{G}_A$ has been completed in $\mathcal{F}$. According to (4.4),

$$\mathbb{P}\{N(A_n) = \boldsymbol{k} \mid \boldsymbol{\xi}\} = \mathbb{P}\{N(A_n) = \boldsymbol{k} \mid \mathcal{G}_A\},$$

and so $\mathbb{E}[V \mid \boldsymbol{\xi}] = \mathbb{E}[V \mid \mathcal{G}_A]$ for any integrable function $V = \varphi(N(A_n))$. Let ζ be any integrable $\mathcal{G}_A$-measurable function. Now $V_m = \mathbb{E}[\zeta | N(A_1), \ldots, N(A_m)]$ is a closed martingale converging a.s. to ζ as $m \to \infty$. And therefore, almost surely

$$\begin{aligned}\zeta &= \lim_m \mathbb{E}[V_m \mid \mathcal{G}_A]\\ &= \lim_m \mathbb{E}[\mathbb{E}[V_m \mid \mathcal{G}_{A_m}] \mid \mathcal{G}_A]\\ &= \lim_m \mathbb{E}[\mathbb{E}[V_m \mid N(A_m)] \mid \mathcal{G}_A]\\ &= \lim_m \mathbb{E}[\mathbb{E}[V_m \mid N(A_m)] \mid \xi]\end{aligned}$$

because V_m and $\{Y(x)\}_{x\in A_m^c}$ are conditionally independent given $N(A_m)$ and the inner conditional expectation is an integrable function of $N(A_m)$. Hence, ζ is the almost sure limit of a sequence of $\sigma(\boldsymbol{\xi})$-measurable functions. Thus $\mathcal{G}_A$ is the completion of $\sigma(\boldsymbol{\xi})$.

Now, assuming that $\sigma(\boldsymbol{\xi})$ is completed in $\mathcal{F}$, we have $\sigma(\boldsymbol{\xi}) = \mathcal{G}_{B\cup A} \subset \mathcal{G}_B$ for any $B \in \mathcal{B}$ (since $B \cup A$ is a regular sequence satisfying the condition (A*)). Hence $\mathcal{G}_{E'} = \bigcap_{B\in\mathcal{B}, B\subset E'} \mathcal{G}_B \supset \sigma(\boldsymbol{\xi})$, while on the other hand, $\mathcal{G}_{E'} \subset \mathcal{G}_A = \sigma(\boldsymbol{\xi})$.

This completes the proof of Corollary 1.

5 Application to an estimation problem

In this section we show an application of the De Finetti Theorem 1 and the strong law of large numbers in Theorem 2 to an estimation problem.

Suppose that ℓ channels flow into a reservoir, where the channel i has an unknown flow of z_i particles per time unit ($t = 1, 2, \ldots$), for each $i = 1, \ldots, \ell$. The problem is to estimate the proportion of particles in the reservoir that come from channel i, that is, the quantities

$$\rho_i = \frac{z_i}{z_1 + \cdots + z_\ell} \quad \text{for } i = 1, \ldots, \ell,$$

with the restriction that we cannot directly take samples from the channels, and instead, sampling from the reservoir is allowed.

The proposed solution consists in adding some "distinguishable" particles of, say, type i to the flow of channel i. Then, by sampling particles from the reservoir and counting the number of such distinguishable particles, we will propose estimators of the ρ_i. The consistency of these estimators will be proved using the De Finetti theorem and the law of large numbers developed in this paper.

We assume that $(z_1, \ldots, z_\ell)$ is an unobservable random vector taking values in $(0, \infty)^\ell$. At each time $t \geq 1$, we add a proportion $f_{i,t} \geq 0$ of particles of type i to the flow of channel i. Therefore, at each time $t \geq 1$, the reservoir receives z_i normal particles (of type 0) and $f_{i,t} z_i$ particles of type i that come from channel i. So, at time $t \geq 1$, the composition of the reservoir is:

- $(z_1 + \cdots + z_\ell)t$ particles of type 0;
- $z_i F_{i,t} = z_i \cdot \sum_{s=1}^{t} f_{i,s}$ particles of type i, for each $1 \leq i \leq \ell$.

The estimation procedure is to pick, at each time $t \geq 1$, a particle from the reservoir. We denote by $Y(t) \in \{0, 1, \ldots, \ell\}$ the type of this sampled particle. Upon observation of the sequence $\{Y(t)\}$, we must estimate the proportions ρ_i. Let us now show that $\{Y(t)\}_{t \geq 1}$ is a random assignment process on the set E of times $t = 1, 2, \ldots$.

Given $B \in \mathcal{B}$ and $\boldsymbol{r} \in R(\#B)$, we have that $\mathbb{P}\{N(B) = \boldsymbol{r}\}$ equals

$$\mathbb{P}\{N(B) = \boldsymbol{r}\} = \mathbb{E}\left[\sum_{B_0, \ldots, B_\ell} \prod_{t \in B_0} \frac{(z_1 + \cdots + z_\ell)t}{\sum_{i=1}^{\ell} z_i (t + F_{i,t})} \cdot \prod_{i=1}^{\ell} \prod_{t \in B_i} \frac{z_i F_{i,t}}{\sum_{i=1}^{\ell} z_i (t + F_{i,t})}\right],$$

where the sum ranges over all partitions B_i of B with $\#B_i = r_i$ for $i = 0, \ldots, \ell$. Hence,

$$\mathbb{P}\{N(B) = \boldsymbol{r}\} = \mathbb{E}\left[\frac{(z_1 + \cdots + z_\ell)^{r_0} \prod_{i=1}^{\ell} z_i^{r_i}}{\prod_{t \in B} \sum_{i=1}^{\ell} z_i (1 + F_{i,t}/t)}\right] \cdot \sum_{B_0, \ldots, B_\ell} \prod_{i=1}^{\ell} \prod_{t \in B_i} F_{i,t}/t.$$

This suggest the following definition of the weight function $\boldsymbol{m}$. Let

$$m_i(t) = F_{i,t}/t \quad \text{for } i = 1, \ldots, \ell, \quad \text{and} \quad m_0(t) = 1$$

for $t \geq 1$. Also, let

$$G(B, \boldsymbol{r}) = \mathbb{E}\left[\frac{(z_1 + \cdots + z_\ell)^{r_0} \prod_{i=1}^{\ell} z_i^{r_i}}{\prod_{t \in B} \sum_{i=1}^{\ell} z_i (1 + F_{i,t}/t)}\right].$$

We have thus shown that

$$\mathbb{P}\{N(B) = \boldsymbol{r}\} = S(B, \boldsymbol{r}) G(B, \boldsymbol{r}).$$

Similarly, if $\{x_1, \ldots, x_k\} \subseteq B$ we have

$$\begin{aligned} &\mathbb{P}\{N(B) = \boldsymbol{r}, Y(x_1) = i_1, \ldots, Y(x_k) = i_k\} \\ &\quad = S\big(B - \{x_1, \ldots, x_k\}, \boldsymbol{r} - (\boldsymbol{e}_{i_1} + \cdots + \boldsymbol{e}_{i_k})\big) \cdot G(B, \boldsymbol{r}) \cdot m_{i_1}(x_1) \cdots m_{i_k}(x_k), \end{aligned}$$

while for $\{y_1, \ldots, y_n\} \cap B = \emptyset$ we have

$$\begin{aligned} \mathbb{P}\{N(B) = \boldsymbol{r}, Y(x_1) = i_1, \ldots, Y(x_k) = i_k, Y(y_1) = j_1, \ldots, Y(y_n) = j_n\} \\ = \ S\big(B - \{x_1, \ldots, x_k\}, \boldsymbol{r} - (\boldsymbol{e}_{i_1} + \cdots + \boldsymbol{e}_{i_k})\big) \\ \cdot\, G\big(B \cup \{y_1, \ldots, y_n\}, \boldsymbol{r} + \boldsymbol{e}_{j_1} + \cdots + \boldsymbol{e}_{j_n}\big) \\ \cdot\, m_{i_1}(x_1) \cdots m_{i_k}(x_k) \cdot m_{j_1}(y_1) \cdots m_{j_n}(y_n). \end{aligned}$$

From the above equalities, it follows easily that the random variables $Y(x_i)$ are conditionally independent of $Y(y_j)$ given $N(B)$. Hence, the condition (C1) is satisfied. It is also straightforward that (C2) is satisfied. Hence, the random variables $\{Y(t)\}_{t\geq 1}$ are a random assignment process with weight function $\boldsymbol{m}$ as defined above.

Now we note that

$$G(B, \boldsymbol{r}) = \mathbb{E}\left[\frac{(1/2)^{r_0} \prod_{i=1}^{\ell} (\rho_i/2)^{r_i}}{\prod_{t \in B} \left(\frac{1}{2} + \sum_{i=1}^{\ell} m_i(t) \frac{\rho_i}{2}\right)}\right].$$

Therefore, letting

$$\boldsymbol{\xi} = \big(1/2, \rho_1/2, \ldots, \rho_\ell/2\big)$$

we have

$$\mathbb{P}\{N(B) = \boldsymbol{r}\} = \int_{\Delta_\ell} \frac{S(B, \boldsymbol{r}) \cdot \xi_0^{r_0} \xi_1^{r_1} \cdots \xi_\ell^{r_\ell}}{\prod_{x \in B} \boldsymbol{\xi} \cdot \boldsymbol{m}(x)} F(d\boldsymbol{\xi}),$$

where F denotes the distribution of $\boldsymbol{\xi}$.

Therefore, the random variables $\{Y(t)\}_{t\geq 1}$ are conditionally independent given the ρ_i. Besides, when condition (A*) is satisfied, then the strong law of large numbers yields almost sure consistent estimators of the relative flows ρ_i. At this point, there are several possible choices of the frequencies $f_{i,t}$.

- If $f_{i,t} = c_i$ for some positive constant c_i and all $t \geq 1$, then $F_{i,t} = c_i t$ (i.e., we add the same proportion of distinguishable particles at each time t). In this case, condition (A*) is satisfied and we have that all the indices $0, 1, \ldots, \ell$ belong to the class C_0. Therefore,

$$\frac{2(N_i(1) + \cdots + N_i(t))}{t} \to \rho_i \quad \text{with probability one as } t \to \infty.$$

  The random variables $\{Y(t)\}$ are exchangeable and they are not independent.
- An interesting choice, which does not require to add so "many" distinguishable particles is $f_{i,1} = c_i > 0$ and $f_{i,t} = 0$ for $t \geq 2$; that is, we add distinguishable particles just at the initial time $t = 1$. We have $F_{i,t} = c_i$. The condition (A*) holds, but now

$$C_0 = \{0\} \quad \text{and} \quad C_1 = \{1, \ldots, \ell\}.$$

  The strong law of large numbers gives

$$\frac{N_0(1) + \cdots + N_0(t)}{t} \to 1 \quad \text{with probability one as } t \to \infty,$$

  that is, the limiting proportion of distinguishable particles in the reservoir vanishes. The consistent estimator is now

$$\frac{N_i(1) + \cdots + N_i(t)}{c_i \log t} \to \rho_i \quad \text{with probability one as } t \to \infty.$$

  In this case, the random variables $\{Y(t)\}$ are neither exchangeable nor independent.

6 Conclusions

In this paper, we have given sufficient conditions for a random assignment process $\{Y(x)\}_{x \in E}$ on $(E, \boldsymbol{m})$ taking values in $L = \{0, 1, \ldots, \ell\}$ to verify a strong law of large numbers and a De Finetti theorem. Namely, it suffices that there exists a sequence $A = \{x_n\}_{n \geq 1} \subset E$ such that the limits

$$\lim_{n \to \infty} \frac{m_j(x_n)}{m_i(x_n)}$$

exist for all $i, j \in L$, and in addition,

$$\sum_{n=1}^{\infty} \frac{m_j(x_n)}{m_i(x_n)} = \infty \quad \text{for all } i, j \in L. \tag{6.1}$$

For the particular case of a random selection process (in the terminology of Vélez and Prieto-Rumeau 2009, 2010, 2011), that is, when $\ell = 1$, the weight function $\boldsymbol{m}$ takes

the particular form

$$\boldsymbol{m}(x) = (m_0(x), m_1(x)) = (1, m(x)) \quad \text{for } x \in E.$$

Then the condition (6.1) becomes

$$\sum_{n=1}^{\infty} m(x_n) = \infty \quad \text{and} \quad \sum_{n=1}^{\infty} 1/m(x_n) = \infty. \tag{6.2}$$

It is shown in Vélez and Prieto-Rumeau (2010) that (6.2) is also a necessary condition for the random selection process to verify a De Finetti theorem. In this sense, an interesting open issue is to know whether the existence of a regular sequence $A \subset E$ that satisfies (6.1) is a necessary condition for the De Finetti theorem.

Finally, a challenging issue—and likely to be fairly complicated—is to generalize the results in the present paper to the case of a random assignment process with a denumerable or continuous set of indices L.

References

Aldous DJ (1981) Representations for partially exchangeable arrays of random variables. J Multivariate Anal 11:581–598

Aldous DJ (1982) On exchangeability and conditional independence. In: Exchangeability in probability and statistics, Rome, 1981. North-Holland, Amsterdam-New York, pp 165–170

Aldous DJ (1985) Exchangeability and related topics. Lecture Notes in Math 1117. Springer, Berlin

Chow YS, Teicher H (1988) Probability theory. Independence, interchangeability, martingales. Springer-Verlag, New York

De Finetti B (1930) Funzione caratteristica di un fenomeno aleatorio. Memorie della R Accademia dei Lincei IV:86–133

De Finetti B (1937) La prévision: ses lois logiques, ses sources subjectives. Ann Inst H Poincaré 7:1–68

Diaconis P, Freedman D (1980) Finite exchangeable sequences. Ann Probab 8:745–764

Diaconis P, Freedman D (1980) De Finetti's theorem for Markov chains. Ann Probab 8:115–130

Diaconis P, Freedman D (1984) Partial exchangeability and sufficiency. In: Statistics: applications and new directions, Calcutta, 1981. Indian Statist Inst, Calcutta, India, pp 205–236

Dovbysh LN, Sudakov VN (1982) Gram-de Finetti matrices. J Soviet Math 24:3047–3054

Dubins LE, Freedman DA (1979) Exchangeable processes need not be mixtures of independent, identically distributed random variables. Z Wahrsch Verw Gebiete 48:115–132

Dynkin EB (1978) Sufficient statistics and extreme points. Ann Probab 6:705–730

Feller W (1966) An introduction to probability theory and its applications, vol II, 2nd edn. John Wiley, New York

Freedman DA (1996) De Finetti's theorem in continuous time. In: Statistics, probability and game theory, IMS Lecture Notes Monogr Ser 30. Inst Math Statist, Hayward, CA, pp 83–98

Fristedt B, Gray L (1997) A modern approach to probability theory. Birkhäuser, Boston

Gnedin AV (1997) The representation of composition structures. Ann Probab 25:1437–1450

Gnedin AV, Pitman J (2004) Regenerative partition structures. Electron J Combin 11(2):R12

Hewitt E, Savage LJ (1955) Symmetric measures on cartesian products. Trans Am Math Soc 80:470–501

Hoover DN (1982) Row–column exchangeability and a generalized model for probability. In: Exchangeability in probability and statistics, Rome, 1981. North-Holland, Amsterdam-New York, pp 281–291

Kallenberg O (1973) A canonical representation of symmetrically distributed random measures. In: Mathematics and statistics (in honour of Harald Bergström on the occasion of his 65th birthday). Department of Math, Chalmers Inst Tech, Göteborg, pp 41–48

Kallenberg O (1975) On symmetrically distributed random measures. Trans Am Math Soc 202:105–121

Kallenberg O (1976) Random measures. Academic Press, London-New York

Kallenberg O (1982) Characterizations and embedding properties in exchangeability. Z Wahrsch Verw Gebiete 60:249–281

Kallenberg O (1982) A dynamical approach to exchangeability. In: Exchangeability in probability and statistics, Rome, 1981. North-Holland, Amsterdam-New York, pp 87–96

Kallenberg O (1989) On the representation theorem for exchangeable arrays. J Multivariate Anal 30: 137–154

Kallenberg O (1992) Symmetries on random arrays and set-indexed processes. J Theoret Probab 5:727–765

Kallenberg O (1997) Foundations of modern probability. Springer-Verlag, New York

Kallenberg O (2005) Probabilistic symmetries and invariance principles. Springer, New York

Kerns GJ, Székely GJ (2006) De Finetti's theorem for abstract finite exchangeable sequences. J Theoret Probab 19:589–608

Kingman JFC (1978) The representation of partition structures. J London Math Soc 18:374–380

Lauritzen SL (1982) Statistical models as extremal families. Aalborg University Press, Aalborg

Loève M (1977) Probability theory I, 4th edn. Springer, New York

Panchenko D (2010) On the Dovbysh–Sudakov representation result. Electron Commun Probab 15:330–338

Pitman JW (1978) An extension of de Finetti's Theorem. Adv Appl Probab 10:268–270

Pitman J (1995) Exchangeable and partially exchangeable random partitions. Probab Theory Related Fields 102:145–158

Vélez R, Prieto-Rumeau T (2008) A De Finetti-type theorem for nonexchangeable finite-valued random variables. J Math Anal Appl 347:407–415

Vélez R, Prieto-Rumeau T (2009) De Finetti's-type results for some families of non identically distributed random variables. Electron J Probab 14:72–86

Vélez R, Prieto-Rumeau T (2010) De Finetti-type theorems for random selection processes. Necessary and sufficient conditions. J Math Anal Appl 365:198–209

Vélez R, Prieto-Rumeau T (2011) De Finetti's type theorems for nonexchangeable 0–1 random variables. Test 20:293–310